Supplemental Material to "Comparison of inferential methods in partially identified models in terms of error in coverage probability"*

Federico A. Bugni
Department of Economics
Duke University
federico.bugni@duke.edu.

September 22, 2014

Abstract

This supplement provides all the necessary results to show that a standardized version of the statistics considered in Bugni (2014) satisfy the same conclusions regarding the rate of convergence of the error in the coverage probability. To establish these findings, we require strengthening the finite moment requirements from finite fourth absolute moments to slightly over sixth absolute moments.

*Thanks to an anonymous referee for suggesting me to explore this extension. Takuya Ura provided excellent research assistance. Any and all errors are my own.

1 Introduction

Bugni (2014) considers the problem of inference in a partially identified moment (in)equality models based on the criterion function approach. According to Assumptions CF.1-CF.2, the sample analogue criterion functions are of the following form

$$Q_n(\theta) = G(\{[\sqrt{n}\bar{m}_{n,j}(\theta)]_-\}_{j=1}^J), \tag{1.1}$$

where G is a non-stochastic function. The fact that G is non-stochastic implies that the sample moment conditions are not allowed to be standardized, i.e., divided by the sample standard deviation. This could be considered an important limitation relative to the existing literature.

The objective of this supplement is to show that the results in Bugni (2014) can be extended to allow for standardized sample moment conditions, i.e., criterion functions of the form

$$Q_n(\theta) = G(\{[\sqrt{n}\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)]_-\}_{j=1}^J), \tag{1.2}$$

where $\hat{\sigma}_{n,j}^2(\theta) \equiv n^{-1}\sum_{i=1}^n (m(W_i,\theta) - \bar{m}_{n,j}(\theta))^2$ is the sample variance of $\{m_j(W_i,\theta)\}_{i=1}^n$ for $j = 1,\ldots,J$.

The strategy to establish these results follows the one used in Bugni (2014). The results in that paper are based on several representation theorems that express a statistic of interest as a well-understood random variable plus an error term, which is shown to converge to zero at a sufficiently fast rate. These type of representation results are obtained for the sample statistic and each of the approximating statistics (bootstrap, asymptotic approximation, subsampling 1, and subsampling 2). This supplement shows that standardization adds a new component to the error term of each of these representation results. As a consequence, all of the results in Bugni (2014) extend for the criterion functions in Eq. (1.2) as long as this additional component converges to zero at an appropriate rate. This supplement shows that occurs, provided that we strengthen the finite moment requirements from finite fourth absolute moments to slightly over sixth absolute moments.

We introduce the following notation in addition to the one used in Bugni (2014). For all $j = 1,\ldots,J$,

$$\begin{aligned}
\sigma_j^2(\theta) &\equiv V[m_j(Z,\theta)] \\
\hat{\sigma}_{n,j}^2(\theta) &\equiv n^{-1}\sum_{i=1}^n (m_j(Z_i,\theta) - \bar{m}_{n,j}(\theta))^2 \\
\tilde{\sigma}_{n,j}^2(\theta) &\equiv n^{-1}\sum_{i=1}^n (m_j(Z_i,\theta) - E[m_j(Z_i,\theta)])^2 \\
v_{n,j}^s(\theta) &\equiv \sqrt{n}(\bar{m}_{n,j}(\theta) - E[m_j(Z,\theta)])/\sigma_j(\theta), \\
\tilde{v}_{n,j}^s(\theta) &\equiv \sqrt{n}(\bar{m}_{n,j}(\theta) - E[m_j(Z,\theta)])/\hat{\sigma}_{n,j}(\theta), \\
v_{n,j}^{s,*}(\theta) &\equiv \sqrt{n}\left(\bar{m}_{n,j}^*(\theta) - \bar{m}_{n,j}(\theta)\right)/\hat{\sigma}_{n,j}(\theta),
\end{aligned}$$

where the superscript "s" in $v_{n,j}^s(\theta)$, $\tilde{v}_{n,j}^s(\theta)$, and $v_{n,j}^{s,*}(\theta)$ denotes that the sample statistic is standardized.

2 Assumptions

In order to develop the results in this supplement, we replace Assumptions A.5, CF.1, and CF.2 in Bugni (2014) by the following alternative assumptions.

Assumption A*.5. $E[||m(Z,\theta)||^{6+\delta}] < \infty$ for some $\delta > 0$.

Assumption CF*.1. The population criterion function is $Q(\theta) = G(\{[E[m_j(Z,\theta)]/\sigma_j(\theta)]_-\}_{j=1}^J)$, where $G : \mathbb{R}_+^J \to \mathbb{R}$ is a non-stochastic and non-negative function that is strictly increasing in every coordinate, weakly convex, continuous, homogeneous of degree $\beta > 0$, and satisfies $G(y) = 0$ if and only if $y = \mathbf{0}_J$.

Assumption CF*.2. The population criterion function is $Q(\theta) = G(\{[E[m_j(Z,\theta)]/\sigma_j(\theta)]_-\}_{j=1}^J)$, where $G : \mathbb{R}_+^J \to \mathbb{R}$ is one of the following two functions: (a) $G(x) = \sum_{j=1}^J \varpi_j x_j$ or (b) $G(x) = \max\{\varpi_j x_j\}_{j=1}^J$, where $\varpi \in \mathbb{R}_+^J$ is an arbitrary vector of positive constants.

Assumption A*.5 strengthens the finite moment requirements from finite fourth absolute moments in Assumption A.5 to slightly over sixth absolute moments. By virtue of Assumptions CF*.1-CF*.2, the properly scaled sample analogue criterion function is now given by Eq. (1.2).

3 Representation results with standardization

This section establishes representation results for the sample statistic and each of the approximating statistics. We note that all the proofs and several intermediate results are collected in the appendix of this supplement. We begin with the representation result for the standardized sample statistic.

Theorem 3.1 (Representation result - Sample statistic)**.** *Assume Assumptions A.1-A.3, CF*.1, and that θ satisfies Assumptions A.4 and A*.5. Let $\rho \equiv rank(V[m(Z,\theta)])$.*

1. *If $\theta \in \partial\Theta_I$, then $Q_n(\theta) = H(\sqrt{n}\bar{Y}_n) + \delta_n$, where*
   (a) *for any $C < \infty$, $P(|\delta_n| > Cn^{-1/2}) = o(n^{-1/2})$.*
   (b) *$\bar{Y}_n : \Omega_n \to \mathbb{R}^\rho$ is a zero mean sample average of n i.i.d. observations from a distribution with non-singular variance-covariance matrix $V = \mathbf{I}_\rho$ and finite fourth moments,*
   (c) *$H : \mathbb{R}^\rho \to \mathbb{R}$ is continuous, non-negative, weakly convex, and homogeneous of degree β. $H(y) = 0$ implies for some non-zero vector $b \in \mathbb{R}^\rho$, $b'y \leq 0$. For any $\mu > 0$, any $|h| \geq \mu > 0$, any $C_2 > 0$, any positive sequence $\{g_n\}_{n\geq 1}$, and any positive sequence $\{\varepsilon_n\}_{n\geq 1}$ with $\varepsilon_n = o(1)$, $\{H^{-1}(\{h\}^{\varepsilon_n}) \cap \|y\| \leq C_2\sqrt{g_n}\} \subseteq \{H^{-1}(\{h\})\}^{\eta_n}$ where $\eta_n = O(\sqrt{g_n}\varepsilon_n)$, uniformly in h.*
   (d) *If we add Assumption CF*.2, then for any $\mu > 0$, any $|h| \geq \mu > 0$ and any sequence $\{\varepsilon_n\}_{n\geq 1}$ with $\varepsilon_n = o(1)$, $\{H^{-1}(\{h\}^{\varepsilon_n})\} \subseteq \{H^{-1}(\{h\})\}^{\gamma_n}$ where $\gamma_n \leq O(\varepsilon_n)$, uniformly in h.*
2. *If $\theta \in Int(\Theta_I)$, then $\liminf\{Q_n(\theta) = 0\}$ a.s.*

As a next step, we consider the standardized version of each of the approximating statistics in Bugni (2014): bootstrap, asymptotic approximation, subsampling 1, and subsampling 2. In each case, one can standardize resampling moment inequalities using: (a) the sample standard deviation or (b) the resampling standard deviation. Both of these options give the same formal results but the latter requires slightly longer arguments. For the sake of brevity, we express our approximation in terms of the first option. In particular:

- For the bootstrap, we replace Bugni (2014, Eq. (3.2)) with

$$\begin{aligned} Q_n^*(\theta) &= G(\{[\sqrt{n}\left(\bar{m}_{n,j}^*(\theta) - \bar{m}_{n,j}(\theta)\right)/\hat{\sigma}_{n,j}(\theta)]_- \times 1[\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta) \leq \tau_n/\sqrt{n}]\}_{j=1}^J) \\ &= G(\{[v_{n,j}^{*;s}(\theta)]_- \times 1[\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta) \leq \tau_n/\sqrt{n}]\}_{j=1}^J), \end{aligned}$$

  where $\bar{m}_n^*(\theta)$ is the sample mean for the bootstrap sample.

- For the asymptotic approximation, we replace Bugni (2014, Eq. (4.2)) with

$$Q_n^{AA}(\theta) = G\left(\left\{\left[\sqrt{n}\sum_{i=1}^{n}\zeta_i\left(m_j(Z_i,\theta)-\bar{m}_{n,j}(\theta)\right)/\hat{\sigma}_{n,j}(\theta)\right]_- \times 1[\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)\leq \tau_n/\sqrt{n}]\right\}_{j=1}^{J}\right),$$

  where $\{\zeta_i\}_{i=1}^n$ is an i.i.d. sample with $\zeta_i \sim N(0,1)$, independent of $\{Z_i\}_{i=1}^n$.

- For subsampling 1, we replace Bugni (2014, Eq. (5.2)) with

$$Q_{b_n,n}^{SS_1}(\theta) = G(\{[\sqrt{b_n}(\bar{m}_{n,b_n,j}^{SS}(\theta)-\bar{m}_{n,j}(\theta))/\hat{\sigma}_{n,j}(\theta)]_- \times 1[\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)\leq \tau_n/\sqrt{n}]\}_{j=1}^{J}),$$

  where $\bar{m}_{n,b_n}^{SS}(\theta)$ is the sample mean for the subsampling sample.

- For subsampling 2, we replace Bugni (2014, Eq. (5.4)) with

$$Q_{b_n,n}^{SS_2}(\theta) = G(\{[\sqrt{b_n}\bar{m}_{n,b_n,j}^{SS}(\theta)/\hat{\sigma}_{n,j}(\theta)]_-\}_{j=1}^{J}).$$

  where $\bar{m}_{n,b_n}^{SS}(\theta)$ is the sample mean for the subsampling sample.

We now establish the analogous representation results for each of these standardized approximation methods. It is relevant to point out that the proofs of each of these build heavily on the corresponding representation results in Bugni (2014).

Theorem 3.2 (Representation result - Bootstrap)**.** *Assume Assumptions A.1-A.3, CF***.1, and that θ satisfies Assumptions A.4-A.5. Let $\rho \equiv rank(V[m(Z,\theta)])$.*

1. *If $\theta \in \partial\Theta_I$ then $Q_n^*(\theta) = H(\sqrt{n}\bar{Y}_n^*) + \delta_n^*$, where*
   (a) *for any $C<\infty$, $P(|\delta_n^*| > Cn^{-1/2}|\mathcal{X}_n) = o(n^{-1/2})$ a.s.*
   (b) *$\{\bar{Y}_n^*|\mathcal{X}_n\} : \Omega_n \to \mathbb{R}^\rho$ is a zero (conditional) mean sample average of n i.i.d. observations from a distribution with a (conditional) variance-covariance matrix $\hat{V}$ which is non-singular a.s. and finite (conditional) fourth moments a.s., and $||\hat{V} - I_p|| \leq O_p(n^{-1/2})$.*
   (c) *$H : \mathbb{R}^\rho \to \mathbb{R}$ is the same function as in Theorem 3.1.*
2. *If $\theta \in Int(\Theta_I)$, then $\liminf\{Q_n^*(\theta) = 0\}$ a.s.*

Theorem 3.3 (Representation result - AA)**.** *Assume Assumptions A.1-A.3 , CF***.1, and that θ satisfies Assumptions A.4-A.5. Let $\rho \equiv rank(V[m(Z,\theta)])$.*

1. *If $\theta \in \partial\Theta_I$ then $Q_n^{AA}(\theta) = H(\sqrt{n}\bar{Y}_n^{AA}) + \delta_n^{AA}$, where*
   (a) *for any $C<\infty$, $P(|\delta_n^{AA}| > Cn^{-1/2}|\mathcal{X}_n) = o(n^{-1/2})$ a.s.*
   (b) *$\{\bar{Y}_n^{AA}|\mathcal{X}_n\} : \Omega_n \to \mathbb{R}^\rho$ is a zero (conditional) mean sample average of n i.i.d. observations from a distribution with a (conditional) variance-covariance matrix $\hat{V}$ which is non-singular a.s. and finite (conditional) fourth moments a.s., and $||\hat{V} - I_p|| \leq O_p(n^{-1/2})$.*
   (c) *$H : \mathbb{R}^\rho \to \mathbb{R}$ is the same function as in Theorem 3.1.*
2. *If $\theta \in Int(\Theta_I)$, then $\liminf\{Q_n^{AA}(\theta) = 0\}$ a.s.*

Theorem 3.4 (Representation result - SS1)**.** *Assume Assumptions A.1-A.3, CF*.1, and that θ satisfies Assumptions A.4-A.5. Let $\rho \equiv rank(V[m(Z,\theta)])$.*

1. *If $\theta \in \partial\Theta_I$ then $Q^{SS_1}_{b_n,n}(\theta) = H(\sqrt{b_n}\bar{Y}^{SS}_{b_n,n}) + \delta^{SS_1}_{b_n,n}$, where*
   (a) $\liminf\{P(\delta^{SS_1}_{b_n,n} = 0|\mathcal{X}_n) \geq 1[||v_n(m_\theta)|| \leq \tau_n]\}$ *and* $\liminf\{\delta^{SS_1}_{b_n,n} = 0\}$*, a.s.*
   (b) *$\{\bar{Y}^{SS}_{b_n,n}|\mathcal{X}_n\} : \Omega_n \to \mathbb{R}^\rho$ is a zero (conditional) mean sample average of b_n observations sampled without replacement from a distribution with a (conditional) variance-covariance matrix $\hat{V}$ which is non-singular a.s. and finite (conditional) fourth moments a.s., and $||\hat{V} - \mathbf{I}_\rho|| \leq O_p(n^{-1/2})$.*
   (c) *$H : \mathbb{R}^\rho \to \mathbb{R}$ is the same function as in Theorem 3.1.*
2. *If $\theta \in Int(\Theta_I)$, then $\liminf\{Q^{SS_1}_{b_n,n}(\theta) = 0\}$ a.s.*

Theorem 3.5 (Representation result - SS2)**.** *Assume Assumptions A.1-A.3, CF*.1, and that θ satisfies Assumptions A.4-A.5. Let $\rho \equiv rank(V[m(Z,\theta)])$.*

1. *If $\theta \in \partial\Theta_I$, then $Q^{SS_2}_{b_n,n}(\theta) = H(\sqrt{b_n}\bar{Y}^{SS}_{b_n,n}) + \delta^{SS_2}_{b_n,n}$, where*
   (a) *for some $C > 0$, $P(|\delta^{SS_2}_{b_n,n}| > C\sqrt{(\ln\ln n)b_n/n}|\mathcal{X}_n) = o(b_n^{-1/2})$ a.s.*
   (b) *$\{\bar{Y}^{SS}_{b_n,n}|\mathcal{X}_n\} : \Omega_n \to \mathbb{R}^\rho$ is a zero (conditional) mean sample average of b_n observations sampled without replacement from a distribution with a (conditional) variance-covariance matrix $\hat{V}$ which is non-singular a.s. and finite (conditional) fourth moments a.s., and $||\hat{V} - \mathbf{I}_\rho|| \leq O_p(n^{-1/2})$.*
   (c) *$H : \mathbb{R}^\rho \to \mathbb{R}$ is the same function as in Theorem 3.1.*
2. *If $\theta \in Int(\Theta_I)$, then $\liminf\{Q^{SS_2}_{b_n,n}(\theta) = 0\}$ a.s.*

Remark 3.1. *Notice that the representation result for the sample statistic (i.e. Theorem 3.1) is the only one of these results that requires slightly more than finite sixth absolute moments in Assumption A*.5. All other representation results in this supplement can be established only using finite fourth absolute moments.*

4 Conclusion

This supplement provides all the necessary results to show that a standardized version of the statistics considered by Bugni (2014) satisfy the same conclusions regarding rate of convergence of the error in the coverage probability. Our strategy is to establish that all the representation results used in Bugni (2014) for non-standardized statistics can also be established for standardized statistics, both for the sample statistic and for all approximating statistics (bootstrap, asymptotic approximation, subsampling 1, and subsampling 2). To establish these result we employ slightly longer formal arguments and we strengthen the finite moment requirements from finite fourth absolute moments to slightly more than finite sixth absolute moments. Using the representation results in this supplement, one can repeat the arguments in Bugni (2014) to establish the exact same rates of convergence of the error in the coverage probability.

Appendix A Appendix

A.1 Proofs of theorems

Proof of Theorem 3.1. Part 1. We show the result by slightly modifying the arguments in Bugni (2014, Part 1, Theorem A.1). Let $\Sigma \equiv V[m(Z,\theta)]$ and, thus, $\rho = rank(V[m(Z,\theta)])$, and let $D \equiv Diag(\Sigma)$ and $\Omega \equiv D^{-1/2}\Sigma D^{-1/2}$. By definition, there are $(J-\rho)$ coordinates of $D^{-1/2}m(Z_i,\theta)$ that can be expressed as a linear combination of the remaining ρ coordinates of $D^{-1/2}m(Z_i,\theta)$ for all $i=1,\ldots,n$ (a.s.). We refer to these ρ coordinates as the "fundamental" coordinates. Without loss of generality, we can rearrange $D^{-1/2}m(Z_i,\theta)$ s.t. the fundamental coordinates are the last ρ ones. This implies that there is a matrix $A \in \mathbb{R}^{(J-\rho)\times\rho}$ s.t.

$$D^{-1/2}m(Z_i,\theta) = [A', \mathbf{I}_\rho]'\{m_j(Z_i,\theta)/\sigma_j(\theta)\}_{j=1}^{\rho} \text{ for all } i=1,\ldots,n,\ a.s. \tag{A.1}$$

Let Ω_ρ denote the variance-covariance matrix of $\{m_j(Z_i,\theta)/\sigma_j(\theta)\}_{j=1}^{\rho}$, which is necessarily positive definite and let $\Omega_\rho^{-1/2}$ denote the inverse of its square root. Then, we define $Y_i \equiv \Omega_\rho^{-1/2}\{(m_j(Z_i,\theta)-E[m_j(Z_i,\theta)])/\sigma_j(\theta)\}_{j=1}^{\rho}$ for all $i=1,\ldots,n$ and the matrix $B \equiv [A',\mathbf{I}_\rho]'\Omega_\rho^{1/2} \in \mathbb{R}^{J\times\rho}$.

By these definitions and by Eq. (A.1), we then conclude that $BY_i = D^{-1/2}(m(Z_i,\theta) - E[m(Z_i,\theta)])$ for all $i=1,\ldots,n$. According to this definition, $v_n^s(\theta) = \sqrt{n}B\bar{Y}_n$, $E[Y_i] = \mathbf{0}_\rho$, $V(Y_i) = \mathbf{I}_p$, and $E[||Y_i||^c] < \infty$ for all $c>0$ s.t. $E[||m(Z,\theta)||^c] < \infty$. Moreover, $\{m(Z_i,\theta)\}_{i=1}^n$ are i.i.d., and so $\{Y_i\}_{i=1}^n$ are also i.i.d.

Let $B_j \in \mathbb{R}^{1\times\rho}$ denote the j^{th} row of B. The function $H(y) : \mathbb{R}^\rho \to \mathbb{R}$ is defined as $H(y) \equiv G(\{[B_j y]_- 1[E[m_j(Z,\theta)]=0]\}_{j=1}^J)$. The same arguments in Bugni (2014, Theorem A.1) can be used to show that this function has all the desired properties. By definition,

$$H(\sqrt{n}\bar{Y}_n) \equiv G(\{[B_j\sqrt{n}\bar{Y}_n]_- 1[E[m_j(Z,\theta)]=0]\}_{j=1}^J)) = G(\{[v_{n,j}^s(\theta)]_- 1[E[m_j(Z,\theta)]=0]\}_{j=1}^J))$$

and $\delta_n \equiv Q_n(\theta) - H(\sqrt{n}\bar{Y}_n)$. In turn, $\delta_n = \delta_{n,1} + \delta_{n,2}$ where

$$\begin{aligned}
\delta_{n,1} &\equiv Q_n(\theta) - G(\{\left[\tilde{v}_{n,j}^s(\theta)\right]_- 1[E[m_j(Z,\theta)]=0]\}_{j=1}^J),\\
\delta_{n,2} &\equiv G(\{\left[\tilde{v}_{n,j}^s(\theta)\right]_- 1[E[m_j(Z,\theta)]=0]\}_{j=1}^J) - G(\{\left[v_{n,j}^s(\theta)\right]_- 1[E[m_j(Z,\theta)]=0]\}_{j=1}^J).
\end{aligned} \tag{A.2}$$

We consider these two terms in separate steps. Step 1 shows that shows that $P(|\delta_{n,1}|>0) = o(n^{-1/2})$ by a slightly modifying the arguments in Bugni (2014, Theorem A.1) and Step 2 shows that $P(|\delta_{n,2}| > Cn^{-1/2}) = o(n^{-1/2})$ for all $C<\infty$. The combination of these two steps completes the proof of this part.

Step 1: Argument for $\delta_{n,1}$. Since $\theta \in \partial\Theta_I \subseteq \Theta_I$ then $E[m(Z,\theta)] \geq \mathbf{0}_J$. Let $S \equiv \{j \in \{1,\ldots,J\} : E[m_j(Z,\theta)] = 0\}$, $\bar{S} \equiv \{1,\ldots,J\}/S$ (which may be empty), $\eta \equiv \min_{j\in\bar{S}} E[m_j(Z,\theta)]/\sigma_j(\theta) > 0$ if $\bar{S} \neq \emptyset$ and $\eta \equiv 0$ if $\bar{S} = \emptyset$. For any $j=1,\ldots,J$, consider the following argument. First, notice that:

$$\begin{aligned}
[\sqrt{n}\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)]_- &= [\sqrt{n}\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)]_- 1[E[m_j(Z,\theta)]=0] + [\sqrt{n}\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)]_- 1[E[m_j(Z,\theta)]>0]\\
&\geq [\tilde{v}_{n,j}(\theta)]_- 1[E[m_j(Z,\theta)]=0].
\end{aligned}$$

Second, notice that:

$$\begin{aligned}
[\sqrt{n}\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)]_- &= [\sqrt{n}\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)]_- 1[E[m_j(Z,\theta)]=0] + [\sqrt{n}\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)]_- 1[E[m_j(Z,\theta)]>0]\\
&= \left\{\begin{array}{c} [\tilde{v}_{n,j}^s(\theta)]_- 1[E[m_j(Z,\theta)]=0] + \\ [v_{n,j}^s(\theta) + \sqrt{n}E_j[m_j(Z,\theta)]/\sigma_j(\theta)]_- 1[E[m_j(Z,\theta)]>0](\sigma_j(\theta)/\hat{\sigma}_{n,j}(\theta)) \end{array}\right\}\\
&\leq [\tilde{v}_{n,j}^s(\theta)]_- 1[E[m_j(Z,\theta)]=0] + [v_{n,j}^s(\theta) + \sqrt{n}\eta]_- 1[E[m_j(Z,\theta)]>0](\sigma_j(\theta)/\hat{\sigma}_{n,j}(\theta)).
\end{aligned}$$

From both of these, we extract the following conclusions. If $\bar{S} = \emptyset$ then $\delta_{n,1} = 0$ and so $P(|\delta_{n,1}|>0) = 0 = o(n^{-1/2})$. If $\bar{S} \neq \emptyset$ then notice that $\left\{\min_{j\in\bar{S}} v_{n,j}^s(\theta) + \sqrt{n}\eta \geq 0\right\} \subseteq \{\delta_{n,1} = 0\}$. Therefore, $P(|\delta_{n,1}|>0) \leq P(\min_{j\in\bar{S}} v_{n,j}^s(\theta) +$

$\sqrt{n}\eta < 0)$. Thus, the proof of this step is completed by showing that:

$$P(\min_{j\in\bar{S}} v^s_{n,j}(\theta) + \sqrt{n}\eta < 0) = o(n^{-1/2}).$$

To show this, notice that for any $\lambda \in (1/4, 1/2)$ we have that:

$$\begin{aligned} P(\min_{j\in\bar{S}} v^s_{n,j}(\theta) + \sqrt{n}\eta < 0) &= P\left(\cup_{j\in\bar{S}}\{v^s_{n,j}(\theta) < -\sqrt{n}\eta\}\right) \leq \sum_{j\in\bar{S}} P\left(|v^s_{n,j}(\theta)| > \sqrt{n}\eta\}\right) \\ &\leq \sum_{j\in\bar{S}}[1[n^\lambda > \sqrt{n}\eta] + P(|v^s_{n,j}(\theta)| > n^\lambda)] = o(n^{-1/2}), \end{aligned}$$

where the rate of convergence follows from $\eta > 0$ ($S_1 \neq \emptyset$) and Lemma A.2.

Step 2: Argument for $\delta_{n,2}$. We show that $P\left(|\delta_{n,2}| > Cn^{-1/2}\right) = o(n^{-1/2})$ for all $C < \infty$. This part of the proof is the only genuinely new part of the argument relative to Bugni (2014, Theorem A.1).

Fix $c > 0$ arbitrarily small so that $\delta > (4+2\delta)2c + 4$ and so $O(n^{1/2-\delta/4+(1+\delta/2)2c}) = o(n^{-1/2})$. Then,

$$\begin{aligned} P(|\delta_{n,2}| > Cn^{-1/2}) &= \left\{ \begin{array}{c} P(\{|\delta_{n,2}| > Cn^{-1/2}\} \cap \{\cap_{j=1}^J\{\{|v^s_{n,j}(\theta)| \leq n^{c/2\beta}\} \cap \{|\hat{\sigma}_{n,j}(\theta) - \sigma_j(\theta)| \leq n^{-1/2-2c}\}\}\}) \\ +P(\{|\delta_{n,2}| > Cn^{-1/2}\} \cap \{\cup_{j=1}^J\{\{|v^s_{n,j}(\theta)| > n^{c/2\beta}\} \cup \{|\hat{\sigma}_{n,j}(\theta) - \sigma_j(\theta)| > n^{-1/2-2c}\}\}\}) \end{array} \right\} \\ &\leq \sum_{j=1}^J P(|v^s_{n,j}(\theta)| > n^{c/2\beta}) + \sum_{j=1}^J P(|\hat{\sigma}_{n,j}(\theta) - \sigma_j(\theta)| > n^{-1/2-2c}), \end{aligned}$$

where the inequality holds by Lemma A.4 for all sufficiently large n. The first sum is $o(n^{-1/2})$ by Lemma A.2 (with $\lambda = c/2\beta$) and the second sum is $o(n^{-1/2})$ by Lemma A.8 (with $\lambda = 2c$). This completes the proof of this step.

Part 2. If $\theta \in Int(\Theta_I)$ then $E[m_j(Z,\theta)] > 0$ for all $j = 1,\ldots,J$. Define $\eta = \min_{j=1,\ldots,J}\{E[m_j(Z,\theta)]/\sigma_j(\theta)\} > 0$. Consider any positive sequence $\{\varepsilon_n\}_{n\geq 1}$ s.t. $\sqrt{\ln\ln n}/\varepsilon_n = o(1)$ and $\varepsilon_n/\sqrt{n} = o(1)$. Suppose that the event $\{\|v^s_n(m_\theta)\| \leq \varepsilon_n\}$ occurs. Then, for all n large enough, $Q_n(\theta) \leq G(\{[(-\varepsilon_n/\sqrt{n}+\eta)\sqrt{n}]_-\}_{j=1}^J) = 0$. Therefore, $\liminf\{\|v^s_n(m_\theta)\| \leq \varepsilon_n\} \subseteq \liminf\{Q_n(\theta) = 0\}$. By the LIL, $P(\liminf\{\|v^s_n(m_\theta)\| \leq \varepsilon_n\}) = 1$ and the result follows. $\square$

Proof of Theorem 3.2. Part 1. We show the results by slightly modifying the arguments in Bugni (2014, Part 1, Theorem A.2). Let the matrices $\Sigma \in \mathbb{R}^{J\times J}$, $D \equiv Diag(\Sigma)$, $A \in \mathbb{R}^{(J-\rho)\times\rho}$, $\Omega_\rho \in \mathbb{R}^{\rho\times\rho}$, and $B \equiv [A', \mathbf{I}_\rho]'\Omega_\rho^{1/2} \in \mathbb{R}^{J\times\rho}$ be defined as in proof of Theorem 3.1. Also, define $\hat{\Sigma} \equiv \hat{V}[m(Z,\theta)]$ and $\hat{D} \equiv Diag(\hat{\Sigma})$.

Since bootstrap samples are constructed from the original random sample, it has to be the case that the coordinates of $\{m(Z_i^*,\theta)\}_{i=1}^n$ can be arranged into the same ρ "fundamental" and $(J-\rho)$ "non-fundamental" coordinates described in Theorem 3.1. In particular, the bootstrap sample satisfies Eq. (A.1). Then, by the argument in Theorem 3.1, we define $Y_i^* \equiv \Omega_\rho^{-1/2}\{(m_j(Z_i^*,\theta) - \bar{m}_{n,j}(\theta))/\hat{\sigma}_j(\theta)\}_{j=1}^\rho$ for all $i = 1,\ldots,n$, which can be shown to satisfy $BY_i^* = \hat{D}^{-1/2}(m(Z_i^*,\theta) - \bar{m}_n(\theta))$ for all $i = 1,\ldots,n$.

According to this definition of $\{Y_i\}_{i=1}^n$, $v_n^{s,*}(\theta) = \sqrt{n}B\bar{Y}_n^*$, $E[Y_i^*|\mathcal{X}_n] = \mathbf{0}_\rho$, $V[Y_i^*|\mathcal{X}_n] = \hat{V}$ where $\hat{V} = \Omega_\rho^{-1/2}\hat{\Omega}_\rho\Omega_\rho^{-1/2}$, where $\hat{\Omega}_\rho$ denotes the sample correlation of $\{m_j(Z,\theta)\}_{j=1}^\rho$. By the SLLN, $\hat{\Sigma} - \Sigma = o(1)$ a.s. which implies that $\hat{D} - D = o(1)$ a.s. and $\hat{\Omega}_\rho - \Omega_\rho = o(1)$ a.s. By this and the CMT, it then follows that $\hat{V} - \mathbf{I}_\rho = o(1)$ a.s., implying that $\hat{V}$ is non-singular, a.s. By a similar argument, the SLLN implies that $E(||Y_i^*||^c|\mathcal{X}_n) < \infty$ for all $c > 0$ s.t. $E[\|m(Z,\theta)\|^c] < \infty$. Moreover, $\{\{Y_i^*\}_{i=1}^n|\mathcal{X}_n\}$ are i.i.d. because $\{\{m(Z_i^*,\theta)\}_{i=1}^n|\mathcal{X}_n\}$ are i.i.d. Finally, if we add Assumption A.5, the CLT and Slutzky's theorem imply that $\sqrt{n}(\hat{\Omega} - \hat{\Omega}) = O_p(1)$ and, so, $||\hat{V} - \mathbf{I}_\rho|| \leq O_p(n^{-1/2})$.

Let $H(y) : \mathbb{R}^\rho \to \mathbb{R}$ be defined as in Theorem 3.1. The same arguments in Bugni (2014, Theorem A.1) can be used to show that this function has all the desired properties. By definition, it then follows that

$$H(\sqrt{n}\bar{Y}_n^*) \equiv G(\{[B_j\sqrt{n}\bar{Y}_n^*]_- 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^J)) = G(\{[v_{n,j}^{*,s}(\theta)]_- 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^J))$$

and $\delta_n^* \equiv Q_n^*(\theta) - H(\sqrt{n}\bar{Y}_n^*)$. To conclude the proof, it suffices to show that this error term satisfies $P(|\delta_n^*| > 0|\mathcal{X}_n) = o(n^{-1/2})$. Since $\theta \in \partial\Theta_I \subseteq \Theta_I$ then $E[m_j(Z,\theta)] \geq 0\ \forall j = 1,\ldots,J$. Let $S \equiv \{j \in \{1,\ldots,J\} : E[m_j(Z,\theta)] > 0\}$,

$\bar{S} \equiv \{1,\ldots,J\}/S$ (which may be empty), $\eta \equiv \min_{j\in\bar{S}} E[m_j(Z,\theta)]/\sigma_j(\theta) > 0$ if $\bar{S} \neq \emptyset$ and $\eta \equiv 0$ if $\bar{S} = \emptyset$, $\sigma_L^2(\theta) \equiv \min_{j=1,\ldots,J} \sigma_j^2(\theta)$, $\sigma_H^2(\theta) \equiv \max_{j=1,\ldots,J} \sigma_j^2(\theta)$, $C_L \equiv (1+\sigma_L^2(\theta)/(2\sigma_H^2))^{-1/2}$, and $C_H \equiv (1-\sigma_L^2(\theta)/(2\sigma_H^2))^{-1/2} > 0$. Consider the following derivation:

$$
\begin{aligned}
&\{\|v_n^s(m_\theta)\| \le \tau_n/C_H\} \cap \{||\hat{\sigma}_n^2(\theta) - \sigma^2(\theta)|| \le \sigma_L^2(\theta)/2\} \\
&\subseteq \cap_{j=1}^J \{\{|\tilde{v}_{n,j}^s(m_\theta)| \le (\sigma_j(\theta)/\hat{\sigma}_{n,j}(\theta))/(\tau_n/C_H)\} \cap \{C_L \le \sigma_j(\theta)/\hat{\sigma}_{n,j}(\theta) \le C_H\}\} \\
&\subseteq \cap_{j=1}^J \{\{\sqrt{n}|(\bar{m}_{n,j}(\theta) - E[m_j(Z,\theta)])|/\hat{\sigma}_{n,j}(\theta) \le \tau_n\} \cap \{C_L \le \sigma_j(\theta)/\hat{\sigma}_{n,j}(\theta) \le C_H\}\} \\
&\subseteq \{\cap_{j\in S}\{\sqrt{n}\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta) \le \tau_n\}\} \cap \{\cap_{j\in\bar{S}}\{\sqrt{n}\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta) \ge C_L\sqrt{n}\eta - \tau_n\}\} \\
&\subseteq \{\cap_{j\in S}\{\sqrt{n}\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta) \le \tau_n\}\} \cap \{\cap_{j\in\bar{S}}\{\sqrt{n}\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta) > \tau_n\}\} = \{Q_n^*(\theta) = 0\},
\end{aligned}
$$

where all inclusions are based on elementary arguments and the last inclusion holds for all sufficiently large n. From here, it then follows that

$$
\begin{aligned}
P(\delta_n^* = 0|\mathcal{X}_n)n^{1/2} &\le (P(\|v_n^s(m_\theta)\| > \tau_n/C_H|\mathcal{X}_n) + P(||\hat{\sigma}_n^2(\theta) - \sigma^2(\theta)|| > \sigma_L^2(\theta)/2|\mathcal{X}_n) - 1)n^{1/2} \\
&= (1[\|v_n^s(m_\theta)\| > \tau_n/C_H] + 1[||\hat{\sigma}_n^2(\theta) - \sigma^2(\theta)|| > \sigma_L^2(\theta)/2])n^{1/2}.
\end{aligned}
$$

where the equality uses the fact that, conditionally on $\mathcal{X}_n$, the events $\{\|v_n^s(m_\theta)\| > \tau_n/C_H\}$ and $\{||\hat{\sigma}_n^2(\theta) - \sigma^2(\theta)|| > \sigma_L^2(\theta)/2\}$ are non-stochastic. To complete the proof, it suffices to show that $P(\liminf\{\|v_n^s(m_\theta)\| > \tau_n/C_H\}) = 1$ and $P(\liminf\{||\hat{\sigma}_n^2(\theta) - \sigma^2(\theta)|| < \sigma_L^2(\theta)/2\}) = 1$. These two results follow from the LIL and the SLLN, respectively.

Part 2. If $\theta \in Int(\Theta_I)$ then $E[m_j(Z,\theta)] > 0$ for all $j = 1,\ldots,J$. Let $\eta \equiv \min_{j=1,\ldots,J} E[m_j(Z,\theta)]/\sigma_j(\theta) > 0$, $\sigma_L^2(\theta) \equiv \min_{j=1,\ldots,J} \sigma_j^2(\theta)$, $\sigma_H^2(\theta) \equiv \max_{j=1,\ldots,J} \sigma_j^2(\theta)$, $C_L \equiv (1+\sigma_L^2(\theta)/(2\sigma_H^2))^{-1/2}$, and $C_H \equiv (1-\sigma_L^2(\theta)/(2\sigma_H^2))^{-1/2} > 0$. By the same argument as in part 1, we have the following derivation:

$$
\begin{aligned}
&\{\|v_n^s(m_\theta)\| \le \tau_n/C_H\} \cap \{||\hat{\sigma}_n^2(\theta) - \sigma^2(\theta)|| \le \sigma_L^2(\theta)/2\} \\
&\subseteq \cap_{j=1}^J \{\{|\tilde{v}_{n,j}^s(m_\theta)| \le (\sigma_j(\theta)/\hat{\sigma}_{n,j}(\theta))/(\tau_n/C_H)\} \cap \{C_L \le \sigma_j(\theta)/\hat{\sigma}_{n,j}(\theta) \le C_H\}\} \\
&\subseteq \cap_{j=1}^J \{\{\sqrt{n}|(\bar{m}_{n,j}(\theta) - E[m_j(Z,\theta)])|/\hat{\sigma}_{n,j}(\theta) \le \tau_n\} \cap \{C_L \le \sigma_j(\theta)/\hat{\sigma}_{n,j}(\theta) \le C_H\}\} \\
&\subseteq \cap_{j=1}^J \{\sqrt{n}\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta) \ge C_L\sqrt{n}\eta - \tau_n\} \\
&\subseteq \cap_{j\in\bar{S}}\{\sqrt{n}\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta) > \tau_n\} = \{Q_n^*(\theta) = 0\},
\end{aligned}
$$

where all inclusions are based on elementary arguments and the last inclusion holds for all sufficiently large n. From here, it then follows that

$$
\begin{aligned}
&\{\liminf\{\|v_n^s(m_\theta)\| \le \tau_n/C_H\}\} \cap \{\liminf\{||\hat{\sigma}_n^2(\theta) - \sigma^2(\theta)|| \le \sigma_L^2(\theta)/2\}\} \\
&= \liminf\{\{\|v_n^s(m_\theta)\| \le \tau_n/C_H\} \cap \{||\hat{\sigma}_n^2(\theta) - \sigma^2(\theta)|| \le \sigma_L^2(\theta)/2\}\} \subseteq \liminf\{Q_n^*(\theta) = 0\}.
\end{aligned}
$$

To complete the proof, it suffices to show that $P(\liminf\{\|v_n^s(m_\theta)\| > \tau_n/C_H\}) = 1$ and $P(\liminf\{||\hat{\sigma}_n^2(\theta) - \sigma^2(\theta)|| < \sigma_L^2(\theta)/2\}) = 1$. These two results follow from the LIL and the SLLN, respectively. □

Proof of Theorem 3.3. This proof follows closely the arguments used to prove Theorem 3.2. The only difference is that we replace $\{Y_i\}_{i=1}^n$ with $\{Y_i^{AA}\}_{i=1}^n$ defined by $Y_i^{AA} \equiv \Omega_\rho^{-1/2}\{\zeta_i(m_j(Z_i^*,\theta) - \bar{m}_{n,j}(\theta))/\hat{\sigma}_j(\theta)\}_{j=1}^\rho$. □

Proof of Theorem 3.4. This proof follows closely the arguments used to prove Theorem 3.2. The only difference is that we replace $\{Y_i\}_{i=1}^n$ with $\{Y_i^{SS}\}_{i=1}^{b_n}$ defined by $Y_i^{SS} \equiv \Omega_\rho^{-1/2}\{(m_j(Z_i^{SS},\theta) - \bar{m}_{n,j}(\theta))/\hat{\sigma}_j(\theta)\}_{j=1}^\rho$. □

Proof of Theorem 3.5. We show the results by slightly modifying the arguments in Bugni (2014, Theorem A.15). Since the structure of this proof is different from the one used to prove Theorems 3.3 or 3.4, we cover the main differences.

Part 1. Define $\delta_{b_n,n}^{SS_2} \equiv Q_{b_n,n}^{SS_2}(\theta) - G(\{[v_{b_n,n}^{s,SS}(m_{j,\theta})]_- 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^J)$, where $v_{b_n,n}^{s,SS}(m_{j,\theta}) \equiv \sqrt{b_n}(\bar{m}_{b_n,n,j}^{SS}(\theta) - \bar{m}_{n,j}(\theta))/\hat{\sigma}_{n,j}(\theta)$. Parts (b)-(c) follow from Theorem 3.4 so we focus on part (a).

For arbitrary $\delta \in (0,1/2)$, let $S \equiv \{j \in \{1,\ldots,J\} : E[m_j(Z,\theta)] = 0\}$, $\bar{S} \equiv \{1,\ldots,J\}/S$, $\eta \equiv \min_{j\in\bar{S}} E[m_j(Z,\theta)]/\sigma_j(\theta) > 0$, $\sigma_L^2(\theta) \equiv \min_{j=1,\ldots,J}\sigma_j^2(\theta)$, $\lambda \equiv \max_{j=1,\ldots,J}(1+\delta)\sqrt{2V[m_j(Z,\theta)]}$, and let A_n be defined as follows:

$$A_n \equiv \{|v_{b_n,n}^{s,SS}(m_{j,\theta})| \leq b_n^{(1-\delta)/2}\}_{j=1}^J \cap \{|v_n^s(m_{j,\theta})| \leq \lambda\sqrt{\ln\ln n}\}_{j=1}^J \cap \{||\hat{\sigma}_n^2(\theta) - \sigma^2(\theta)|| \leq \sigma_L^2(\theta)/2\}\}. \tag{A.3}$$

For any $C > 0$, notice that:

$$\sqrt{b_n}P(|\delta_{b_n,n}^{SS_2}| > C\sqrt{(\ln\ln n)b_n/n}|\mathcal{X}_n) \leq \sqrt{b_n}P(\{A_n\}^c|\mathcal{X}_n) + \sqrt{b_n}P(\{|\delta_{b_n,n}^{SS_2}| > C\sqrt{(\ln\ln n)b_n/n}\} \cap A_n|\mathcal{X}_n). \tag{A.4}$$

The proof is completed by showing that the two terms on the RHS of Eq. (A.4) are $o(1)$, a.s.

We begin with the first term in the RHS of Eq. (A.4). Consider the following derivation:

$$\begin{aligned}
\sqrt{b_n}P(\{A_n\}^c|\mathcal{X}_n) &= \left\{\begin{array}{c} \sum_{j=1}^J \sqrt{b_n}P(|v_{b_n,n}^{s,SS}(m_{j,\theta})| > b_n^{(1-\delta)/2}|\mathcal{X}_n) + \sum_{j=1}^J \sqrt{b_n}P(|v_n^s(m_{j,\theta})| > \lambda\sqrt{\ln\ln n}|\mathcal{X}_n) + \\ \sum_{j=1}^J \sqrt{b_n}P(|\hat{\sigma}_{n,j}^2(\theta) - \sigma_j^2(\theta)| > \sigma_L^2(\theta)/2|\mathcal{X}_n) \end{array}\right\} \\
&= \left\{\begin{array}{c} \sum_{j=1}^J \sqrt{b_n}P(|v_{b_n,n}^{s,SS}(m_{j,\theta})| > b_n^{(1-\delta)/2}|\mathcal{X}_n) + \sum_{j=1}^J \sqrt{b_n}1(|v_n^s(m_{j,\theta})| > \lambda\sqrt{\ln\ln n}) + \\ \sum_{j=1}^J \sqrt{b_n}1(|\hat{\sigma}_{n,j}^2(\theta) - \sigma_j^2(\theta)| > \sigma_L^2(\theta)/2), \end{array}\right\},
\end{aligned}$$

where the first equality follows from elementary arguments and the second equality follows from the fact that $\{v_n(m_{j,\theta})|\mathcal{X}_n\}$ is deterministic. Fix $j = 1,\ldots,J$ and $\varepsilon > 0$ arbitrarily. By the LIL, $\liminf\{|v_n(m_{j,\theta})| \leq \lambda\sqrt{\ln\ln n}\}_{j=1}^J$ a.s. and so $P(\lim\sqrt{b_n}1(|v_n(m_{j,\theta})| > \lambda\sqrt{\ln\ln n}) = 0) = 1$. By the SLLN, $\liminf\{|\hat{\sigma}_{n,j}^2(\theta) - \sigma_j^2(\theta)| \leq \sigma_L^2(\theta)/2\}$ a.s. and so $P(\lim\sqrt{b_n}1(|\hat{\sigma}_{n,j}^2(\theta) - \sigma_j^2(\theta)| > \sigma_L^2(\theta)/2) = 0) = 1$. Next, consider the following derivation:

$$\begin{aligned}
&P(\liminf\{P(|v_{b_n,n}^{s,SS}(m_{j,\theta})| > b_n^{(1-\delta)/2}|\mathcal{X}_n)\sqrt{b_n} \leq \varepsilon\}) \\
&\geq P\left(\begin{array}{c} \liminf\{P(|v_{b_n,n}^{s,SS}(m_{j,\theta})| > b_n^{(1-\delta)/2}|\mathcal{X}_n)b_n^{(\delta-1)} \leq (1+\varepsilon)\} \\ \cap\liminf\{b_n^{(\delta-1/2)}(1+\varepsilon) \leq \varepsilon\} \end{array}\right) \\
&\geq \left\{\begin{array}{c} P(\liminf\{P(|v_{b_n,n}^{s,SS}(m_{j,\theta})| > b_n^{(1-\delta)/2}|\mathcal{X}_n)b_n^{(\delta-1)} \leq (1+\varepsilon)\}) \\ +P(\liminf\{b_n^{(\delta-1/2)}(1+\varepsilon) \leq \varepsilon\}) - 1 \end{array}\right\} \\
&= P(\liminf\{P(|v_{b_n,n}^{s,SS}(m_{j,\theta})| > b_n^{(1-\delta)/2}|\mathcal{X}_n)b_n^{(\delta-1)} \leq (1+\varepsilon)\}) \leq P(\liminf\{1 \leq 1+\varepsilon\}) = 1,
\end{aligned}$$

where the two first two inequalities follow from elementary arguments, the following equality follows from $b_n^{(\delta-1/2)} = o(1)$, and the third inequality is shown later in Eq. (A.5).

To complete this argument, consider the following derivation:

$$\begin{aligned}
&E(v_{b_n,n}^{s,SS}(m_{j,\theta})^2|\mathcal{X}_n) \\
&= \left(\begin{array}{c} q_n^{-1}\sum_{s=1}^{q_n} b_n^{-1}\sum_{i=1}^{b_n}(m_j(X_{i,s}^{SS},\theta) - \bar{m}_{n,j}(\theta))^2/\hat{\sigma}_{n,j}^2(\theta) + \\ q_n^{-1}\sum_{s=1}^{q_n} b_n^{-1}\sum_{a=1}^{b_n}\sum_{b=1,b\neq a}^{b_n}(m_j(X_{a,s}^{SS},\theta) - \bar{m}_{n,j}(\theta))(m_j(X_{b,s}^{SS},\theta) - \bar{m}_{n,j}(\theta))/\hat{\sigma}_{n,j}^2(\theta) \end{array}\right) \\
&= \left(\begin{array}{c} n^{-1}\sum_{i=1}^{n}(m_j(X_i,\theta) - \bar{m}_{n,j}(\theta))^2/\hat{\sigma}_{n,j}^2(\theta) + \\ +2n^{-1}(n-1)^{-1}\sum_{a=1}^{n}\sum_{b=1,b\neq a}^{n}(m_j(X_a,\theta) - \bar{m}_{n,j}(\theta))(m_j(X_b,\theta) - \bar{m}_{n,j}(\theta))/\hat{\sigma}_{n,j}^2(\theta) \end{array}\right) \leq 1,
\end{aligned}$$

where the first equality holds by expanding squares, the following equality holds by the fact that we are sampling without replacement, and the final inequality holds by verifying that, in the previous line, the first term equals one and the second term is non-positive by the negative associated produced by sampling without replacement (see Joag-Dev and Proschan (1983, Section 3.2(a))). Chebyshev's inequality then implies that

$$1 \geq E(v_{b_n,n}^{s,SS}(m_{j,\theta})^2|\mathcal{X}_n) \geq P(|v_{b_n,n}^{s,SS}(m_{j,\theta})| > b_n^{(1-\delta)/2}|\mathcal{X}_n)b_n^{(1-\delta)}, \tag{A.5}$$

which completes the proof for the first term on the RHS of Eq. (A.4).

We now consider the second term RHS of Eq. (A.4). Condition on $\mathcal{X}_n$ and assume that A_n occurs. Then,

$$\begin{aligned}
Q^{SS_2}_{b_n,n}(\theta) &= G(\{[v^{s,SS}_{b_n,n}(m_{j,\theta}) + \sqrt{b_n/n}v^s_n(m_{j,\theta}) + \sqrt{b_n}E[m_j(Z,\theta)]/\hat{\sigma}_{n,j}(\theta)]_-\}^J_{j=1}) \quad (A.6)\\
&\leq G(\{[v^{s,SS}_{b_n,n}(m_{j,\theta}) - \lambda\sqrt{(\ln\ln n)b_n/n} + \sqrt{b_n}E[m_j(Z,\theta)]/\hat{\sigma}_{n,j}(\theta)]_-\}^J_{j=1})\\
&\leq G\left(\left\{\begin{array}{c} [v^{s,SS}_{b_n,n}(m_{j,\theta}) - \lambda\sqrt{(\ln\ln n)b_n/n}]_- 1[E[m_j(Z,\theta)]=0] + \\ [-b_n^{1/2-\delta} - \lambda\sqrt{(\ln\ln n)b_n/n} + \sqrt{b_n}2\eta]_- 1[E[m_j(Z,\theta)]>0] \end{array}\right\}^J_{j=1}\right)\\
&= G(\{[v^{s,SS}_{b_n,n}(m_{j,\theta}) - \lambda\sqrt{(\ln\ln n)b_n/n}]_- 1[E[m_j(Z,\theta)]=0]\}^J_{j=1})\\
&\leq G(\{[v^{s,SS}_{b_n,n}(m_{j,\theta})]_- 1[E[m_j(Z,\theta)]=0]\}^J_{j=1} + \lambda\sqrt{(\ln\ln n)b_n/n}1_J), \quad (A.7)
\end{aligned}$$

where the first equality is elementary, the second inequality holds because G is increasing and A_n implies that $v^s_n(m_{j,\theta}) \geq -\lambda\sqrt{\ln\ln n}$, the second inequality follows from the fact that A_n implies that $v^{s,SS}_{b_n,n}(m_{j,\theta}) > -b_n^{(1-\delta)/2}$, the next equality follows for all n large enough as $\eta, \lambda, \delta > 0$, and the final inequality holds by elementary arguments. By using similar arguments, we can establish an analogous lower bound for $Q^{s,SS_2}_{b_n,n}(\theta)$. As a consequence, if A_n occurs,

$$Q^{SS_2}_{b_n,n}(\theta) \in \left[\begin{array}{c} G(\{[v^{s,SS}_{b_n,n}(m_{j,\theta})]_- 1[E[m_j(Z,\theta)]=0]\}^J_{j=1} - \lambda\sqrt{(\ln\ln n)b_n/n}1_J), \\ G(\{[v^{s,SS}_{b_n,n}(m_{j,\theta})]_- 1[E[m_j(Z,\theta)]=0]\}^J_{j=1} + \lambda\sqrt{(\ln\ln n)b_n/n}1_J) \end{array}\right]$$

By Assumption CF.2, $\forall x \in \mathbb{R}^J$ and $\forall\varepsilon > 0$, $\exists D > 0$ s.t. $|G(x+\varepsilon) - G(x)| \leq D||\varepsilon||$. If we set $C = D\lambda > 0$,

$$\begin{aligned}
|\delta^{SS_2}_{b_n,n}| &= |Q^{SS_2}_{b_n,n}(\theta) - G(\{[v^{s,SS}_{b_n,n}(m_{j,\theta})]_- 1[E[m_j(Z,\theta)]=0]\}^J_{j=1})|\\
&\leq \max_{r\in\{-1,1\}}\left\{\left|\begin{array}{c} G(\{[v^{s,SS}_{b_n,n}(m_{j,\theta})]_- 1[E[m_j(Z,\theta)]=0]\}^J_{j=1} + r\lambda\sqrt{(\ln\ln n)b_n/n}1_J) \\ -G(\{[v^{s,SS}_{b_n,n}(m_{j,\theta})]_- 1[E[m_j(Z,\theta)]=0]\}^J_{j=1}) \end{array}\right|\right\}\\
&\leq D\|\lambda\sqrt{(\ln\ln n)b_n/n} \times 1_J\| = D\lambda\sqrt{(\ln\ln n)b_n/n}.
\end{aligned}$$

This completes the argument for the second term on the RHS of Eq. (A.4).

Part 2. Since $Q^{SS_2}_{b_n,n}(\theta) \geq 0$, it suffices to show that $\sqrt{b_n}P(Q^{SS_2}_{b_n,n}(\theta) > 0|\mathcal{X}_n) = o(1)$. For arbitrary $\delta \in (0, 1/2)$, let $\eta \equiv \min_{j=1,\dots,J} E[m_j(Z,\theta)]/\sigma_j(\theta) > 0$ and let A_n be defined as in Eq. (A.3). By elementary arguments

$$\sqrt{b_n}P(Q^{SS_2}_{b_n,n}(\theta) > 0|\mathcal{X}_n) \leq \sqrt{b_n}P(\{A_n\}^c|\mathcal{X}_n) + \sqrt{b_n}P(\{Q^{SS_2}_{b_n,n}(\theta) > 0\} \cap A_n|\mathcal{X}_n).$$

We can now repeat arguments used in part 1 to argue that both terms on the RHS are $o(1)$, a.s. On the one hand, the same argument as in part 1 implies that $\sqrt{b_n}P(\{A_n\}^c|\mathcal{X}_n) = o(1)$ a.s. On the other hand, if A_n occurs, the argument used in Eq. (A.6) implies that $Q^{SS_2}_{b_n,n}(\theta) \leq G(\{[-b_n^{1/2-\delta} - \lambda\sqrt{(\ln\ln n)b_n/n} + \eta 2\sqrt{b_n}]_-\}^J_{j=1})$. Since $\eta, \delta > 0$, the RHS expression is equal to zero for all n large enough. This then implies that $\sqrt{b_n}P(\{Q^{SS_2}_{b_n,n}(\theta) > 0\} \cap A_n|\mathcal{X}_n) = o(1)$, completing the proof. □

A.2 Proofs of intermediate results

Lemma A.1. *Assume Assumption A.1 and that $\theta \in \Theta$ satisfies $E[||m(Z,\theta)||^{2+\psi}] < \infty$ for some $\psi > 0$. For all $j = 1, \dots, J$,*

$$\begin{aligned}
E\left[|m_j(Z_i,\theta) - E[m_j(Z_i,\theta)]|^{2+\psi}\right] &< \infty,\\
E\left[\left|(m_j(Z_i,\theta) - E[m_j(Z_i,\theta)])^2 - \sigma^2_j(\theta)\right|^{1+\psi/2}\right] &< \infty.
\end{aligned}$$

Proof. Fix $j = 1, \dots, J$ arbitrarily and let $\sigma^2 \equiv \sigma^2_j(\theta)$ and $M_i \equiv m_j(Z_i,\theta)$ for all $i = 1, \dots, n$.

By Assumption A.5 and Holder's inequality, $E[|M_i|]^{2+\psi} \leq E[|M_i|^{2+\psi}] < \infty$. Another application of Holder's inequality yields $|M_i - E[M_i]|^{2+\psi} \leq |M_i|^{2+\psi} + |E[M_i]|^{2+\psi}$ and, so, $E|M_i - E[M_i]|^{2+\psi} \leq E|M_i|^{2+\psi} + |E[M_i]|^{2+\psi} < \infty$. This proves the first result.

Define $Y_i \equiv (M_i - E[M_i])^2/\sigma^2 - 1$ and $\phi \equiv \psi/2$. By definition,

$$E[|(M_i - E[M_i])^2 - \sigma^2|^{1+\phi}] = \sigma^{2+\phi} E\left[\left|\frac{(M_i - E[M_i])^2}{\sigma^2} - 1\right|^{1+\phi}\right] = \sigma^{2+\phi} E[|Y_i|^{1+\phi}].$$

So, it suffices to show that $E[|Y_i|^{1+\phi}] < \infty$. For the remainder of the proof let $\beta > 1$ be arbitrarily chosen. Notice that

$$E[|Y_i|^{1+\phi}] = E[|Y_i|^{1+\phi} 1[|Y_i| > \beta]] + E[|Y_i|^{1+\phi} 1[|Y_i| \leq \beta]] \leq E[|Y_i|^{1+\phi} 1[|Y_i| > \beta]] + \beta^{1+\phi}.$$

So, it suffices to show that $E[|Y_i|^{1+\phi} 1[|Y_i| > \beta]] < \infty$. Since $\beta > 1$, it follows that

$$\{|Y_i| > \beta\} \subseteq \left\{\left\{\left|\frac{(M_i - E[M_i])^2}{\sigma^2}\right| > \beta + 1\right\} \cap \left\{|Y_i| \leq \left|\frac{(M_i - E[M_i])^2}{\sigma^2}\right|\right\}\right\}$$

and, therefore,

$$\begin{aligned} E[|Y_i|^{1+\phi} 1[|Y_i| > \beta]] &\leq E\left[\left|\frac{M_i - E[M_i]}{\sigma}\right|^{2+\psi} 1\left[\frac{(M_i - E[M_i])^2}{\sigma^2} > \beta + 1\right]\right] \\ &\leq E\left[\left|\frac{M_i - E[M_i]}{\sigma}\right|^{2+\psi}\right] = \frac{1}{\sigma^{2+\psi}} E[|M_i - E[M_i]|^{2+\psi}] < \infty. \end{aligned}$$

This proves the second result and completes the proof. □

Remark A.1. *Lemma A.1 can be used with $\psi = 2$ under Assumption A.5 or with $\psi > 4$ under Assumption $A^\star.5$.*

Lemma A.2. *Assume Assumption A.1 and that $\theta \in \Theta$ satisfies Assumption A.5. For all $j = 1, \ldots, J$ and $c > 0$,*

$$P[|v^s_{n,j}(\theta)| > n^c] = o(n^{-1/2}).$$

Proof. For any $\psi > 0$, consider the following argument based on Chebyshev's inequality:

$$\begin{aligned} P[|v^s_{n,j}(\theta)| > n^c] n^{(2+\psi)c} &= P[|v^s_{n,j}(\theta)|^{2+\psi} > n^{(2+\psi)c}] n^{(2+\psi)c} \\ &\leq E[|v^s_{n,j}(\theta)|^{2+\psi}] \\ &= E\left[\left|n^{-1/2} \sum_{i=1}^n \frac{[m_j(Z,\theta) - E[m_j(Z,\theta)]]}{\sigma_j(\theta)}\right|^{2+\psi}\right] \\ &= \sigma_j^{-(2+\psi)}(\theta) n^{-1-\psi/2} E\left[\left|\sum_{i=1}^n [m_j(Z,\theta) - E[m_j(Z,\theta)]]\right|^{2+\psi}\right] \\ &\leq \sigma_j^{-(2+\psi)}(\theta) n^{-1-\psi/2} E\left[\left(\sum_{i=1}^n |m_j(Z,\theta) - E[m_j(Z,\theta)]|^{2+\psi}\right)\right] \\ &\leq \sigma_j^{-(2+\psi)}(\theta) n^{-\psi/2} E\left[|m_j(Z,\theta) - E[m_j(Z,\theta)]|^{2+\psi}\right]. \end{aligned}$$

From this, Lemma A.1 with $\psi = 2$, and Assumption A.5, we conclude that

$$P[|v^s_{n,j}(\theta)| > n^c] = O(n^{-\psi/2-(2+\psi)c}) = o(n^{-1/2}),$$

where we have used that $c > 0$ and $\psi \geq 1$. □

Lemma A.3. *Assume Assumption CF⋆.1 and that $\theta \in \Theta_I$ satisfies Assumption A.4. Then, $\delta_{n,2}$ in Eq.* (A.2) *satisfies*

$$|\delta_{n,2}| \leq \left[\left(1 - \tfrac{\max_{j=1,\dots,J} |\hat{\sigma}_{n,j}(\theta) - \sigma_j(\theta)|}{\min_{j=1,\dots,J} \sigma_j(\theta)}\right)^{-\beta} - 1 \right] G(\{[v_{n,j}(\theta)]_{-} 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^{J}),$$

Proof. For each $j = 1,\dots,J$, define $v^{L}_{n,j}(\theta) \equiv \min\{\tilde{v}^{s}_{n,j}(\theta), v^{s}_{n,j}(\theta)\}$ and $v^{H}_{n,j}(\theta) \equiv \max\{\tilde{v}^{s}_{n,j}(\theta), v^{s}_{n,j}(\theta)\}$. Since $\theta \in \Theta_I$, $E[m(Z,\theta)] \geq \mathbf{0}_J$. Define $S_{0,-}(\theta) \equiv \{j = 1,\dots,J : E[m_j(Z,\theta)] = 0 \cap v^{L}_{n,j}(\theta) < 0\}$ (which may be empty).

First, consider the case when $S_{0,-}(\theta) = \emptyset$. In this case, either $E[m_j(Z,\theta)] > 0$ or $0 \leq v^{L}_{n,j}(\theta) \leq v^{H}_{n,j}(\theta)$ for all $j = 1,\dots,J$. Therefore,

$$\{[\tilde{v}^{s}_{n,j}(\theta)]_{-} 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^{J} = \{[v^{s}_{n,j}(\theta)]_{-} 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^{J} = \mathbf{0}_J,$$

and, thus, $\delta_{n,2} = 0$, and the statement holds.

Second, consider the case when $S_{0,-}(\theta) \neq \emptyset$. Define $\alpha \equiv \max_{j \in S_{0,-}(\theta)} \{v^{L}_{n,j}(\theta)/v^{H}_{n,j}(\theta)\} \geq 1$. Now consider the following derivation:

$$\begin{aligned}
\alpha^{\beta} G(\{[v^{H}_{n,j}(\theta)]_{-} 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^{J}) &= G(\{[\alpha v^{H}_{n,j}(\theta)]_{-} 1[E[m_j(Z,\theta)] = 0]\}_{j \in S_{0,-}(\theta)}, \{0\}_{j \notin S_{0,-}(\theta)}) \\
&\geq G(\{[v^{L}_{n,j}(\theta)]_{-} 1[E[m_j(Z,\theta)] = 0]\}_{j \in S_{0,-}(\theta)}, \{0\}_{j \notin S_{0,-}(\theta)}) \\
&= G(\{[v^{L}_{n,j}(\theta)]_{-} 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^{J})
\end{aligned}$$

where the first equality follows from homogeneity of degree β and the fact that $[\alpha v^{H}_{n,j}(\theta)]_{-} 1[E[m_j(Z,\theta)] = 0] = 0$ for all $j \notin S_{0,-}(\theta)$, the first inequality follows from the monotonicity of G and $\alpha v^{H}_{n,j}(\theta) < \{v^{L}_{n,j}(\theta)/v^{H}_{n,j}(\theta)\} v^{H}_{n,j}(\theta) = v^{L}_{n,j}(\theta)$ for all $j \in S_{0,-}(\theta)$, and the final equality follows from the fact that $[v^{L}_{n,j}(\theta)]_{-} 1[E[m_j(Z,\theta)] = 0] = 0$ for all $j \notin S_{0,-}(\theta)$. From here, consider the following derivation:

$$\begin{aligned}
|\delta_{n,2}| &\leq G(\{[v^{L}_{n,j}(\theta)]_{-} 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^{J}) - G(\{[v^{H}_{n,j}(\theta)]_{-} 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^{J}) \\
&\leq (\alpha^{\beta} - 1) G(\{[v^{H}_{n,j}(\theta)]_{-} 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^{J}) \\
&\leq \left[\left(\max_{j=1,\dots,J} \left\{ \max\left\{ \frac{\hat{\sigma}_{n,j}(\theta)}{\sigma_j(\theta)}, \frac{\sigma_j(\theta)}{\hat{\sigma}_{n,j}(\theta)} \right\} \right\} \right)^{\beta} - 1 \right] G(\{[v^{s}_{n,j}(\theta)]_{-} 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^{J}) \\
&\leq \left\{ \begin{array}{c} \left[\left(\max_{j=1,\dots,J} \left\{ \max\left\{ \frac{|\hat{\sigma}_{n,j}(\theta) - \sigma_j(\theta)|}{\sigma_L(\theta)} + 1, \frac{1}{1 - \frac{|\hat{\sigma}_{n,j}(\theta) - \sigma_j(\theta)|}{\sigma_L(\theta)}} \right\} \right\} \right)^{\beta} - 1 \right] \\ \times G(\{[v^{s}_{n,j}(\theta)]_{-} 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^{J}) \end{array} \right\} \\
&\leq \left[\left(1 - \frac{\max_{j=1,\dots,J} |\hat{\sigma}_{n,j}(\theta) - \sigma_j(\theta)|}{\sigma_L(\theta)}\right)^{-\beta} - 1 \right] G(\{[v^{s}_{n,j}(\theta)]_{-} 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^{J}),
\end{aligned}$$

where all inequalities are elementary and based on the definition of α and the monotonicity of G. □

Lemma A.4. *Assume Assumption CF⋆.1 and that $\theta \in \Theta_I$ satisfies Assumption A.4. Then, $\forall c > 0$, $\exists N \in \mathbb{N}$ s.t. $\forall n \geq N$, $\delta_{n,2}$ in Eq.* (A.2) *satisfies*

$$\left\{ \left\{ \max_{j=1,\dots,J} |\hat{\sigma}_{n,j}(\theta) - \sigma_j(\theta)| \leq n^{-1/2-2c} \right\} \cap \left\{ \max_{j=1,\dots,k} |v^{s}_{n,j}(\theta)| \leq n^{c/2\beta} \right\} \right\} \subseteq \left\{ |\delta_{n,2}| \leq n^{-1/2-c} \right\}.$$

Proof. First, assume that $\max_{j=1,\dots,J} |\hat{\sigma}_{n,j}(\theta) - \sigma_j(\theta)| \leq n^{-1/2-2c}$. This implies that $\max_{j=1,\dots,J} |\hat{\sigma}_{n,j}(\theta) - \sigma_j(\theta)| / \min_{j=1,\dots,J} \sigma_j(\theta) \leq 1/2$ for all n sufficiently large. For $x \in [0, 1/2]$ consider the function $f(x) = (1-x)^{-\beta} - 1$. By the intermediate value theorem, there is $\tilde{x} \in [0, 1/2]$ s.t. $f(x) = f(0) + f'(\tilde{x})x = \beta(1-\tilde{x})^{-(\beta+1)}x$ and, so, $|f(x)| < \beta 2^{(\beta+1)} x$. Therefore,

$$\left[\left(1 - \frac{\max_{j=1,\dots,J} |\hat{\sigma}_{n,j}(\theta) - \sigma_j(\theta)|}{\min_{j=1,\dots,J} \sigma_j(\theta)}\right)^{-\beta} - 1 \right] \leq \beta 2^{(\beta+1)} \max_{j=1,\dots,J} |\hat{\sigma}_{n,j}(\theta) - \sigma_j(\theta)| \leq \beta 2^{(\beta+1)} n^{-1/2-2c}.$$

Second, assume that $\max_{j=1,\ldots,J}|v_{n,j}(\theta)| \leq n^{c/2\beta}$. If we combine this with the monotonicity and the homogeneity of G we deduce that

$$G(\{[v^s_{n,j}(\theta)]_{-}1[E[m_j(Z,\theta)]=0]\}_{j=1}^J) \leq n^{c/2}G(\{1\}_{j=1}^J).$$

By combining these two steps with Lemma A.3, we conclude that $\max_{j=1,\ldots,J}|\hat{\sigma}_{n,j}(\theta)-\sigma_j(\theta)| \leq n^{-1/2-2c}$ and $\max_{j=1,\ldots,J}|v^s_{n,j}(\theta)| \leq n^{c/2\beta}$ imply that, for all n sufficiently large,

$$|\delta_{n,2}| \leq \beta 2^{(\beta+1)}n^{-1/2-2c}n^{c/2}G(\{1\}_{j=1}^J) \leq \beta 2^{(\beta+1)}G(\{1\}_{j=1}^J)n^{-1/2-1.5c}.$$

The RHS is less than $n^{-1/2-c}$ for all n sufficiently large, and this completes the proof. $\square$

Lemma A.5. *Assume Assumption A.1 and that $\theta \in \Theta$ satisfies $E[||m(Z,\theta)||^{2+\psi}] < \infty$ for some $\psi > 0$. For all $j = 1,\ldots,J$ and any sequence $\{a_n\}_{n\geq 1} = o(1)$,*

$$P(|\tilde{\sigma}_{n,j}(\theta)-\sigma_j(\theta)| > \sigma a_n)(3\sigma_j(\theta)a_n)^{1+\psi/2} \leq n^{-\psi/2}E[|(m_j(Z_i,\theta)-E[m_j(Z_i,\theta)])^2-\sigma_j^2(\theta)|^{1+\psi/2}].$$

Proof. Consider the following argument:

$$\begin{aligned}
P\left(|\tilde{\sigma}_{n,j}(\theta)-\sigma_j(\theta)| > \sigma_j(\theta)a_n\right) &= P\left(\{\tilde{\sigma}_{n,j}(\theta)-\sigma_j(\theta) > \sigma_j(\theta)a_n\} \cup \{\tilde{\sigma}_{n,j}(\theta)-\sigma_j(\theta) < -\sigma_j(\theta)a_n\}\right) \\
&= P\left(\{\tilde{\sigma}_{n,j}(\theta) > \sigma_j(\theta)\left(1+a_n\right)\} \cup \{\tilde{\sigma}_{n,j}(\theta) < \sigma_j(\theta)\left(1-a_n\right)\}\right) \\
&= P\left(\begin{array}{c} \left\{(\tilde{\sigma}^2_{n,j}(\theta)-\sigma^2_j(\theta)) > \sigma^2_j(\theta)(2a_n+a_n^2)\right\} \cup \\ \left\{(\tilde{\sigma}^2_{n,j}(\theta)-\sigma^2_j(\theta)) < \sigma^2_j(\theta)(-2a_n+a_n^2)\right\} \end{array}\right) \\
&\leq \left\{\begin{array}{c} P\left((\tilde{\sigma}^2_{n,j}(\theta)-\sigma^2_j(\theta)) > \sigma^2_j(\theta)(2a_n+a_n^2)\right) \\ +P\left((\tilde{\sigma}^2_{n,j}(\theta)-\sigma^2_j(\theta)) < \sigma^2_j(\theta)(-2a_n+a_n^2)\right) \end{array}\right\} \\
&\leq P\left((\tilde{\sigma}^2_{n,j}(\theta)-\sigma^2_j(\theta)) > 3\sigma^2_j(\theta)a_n\right) + P\left((\tilde{\sigma}^2_{n,j}(\theta)-\sigma^2_j(\theta)) > -3\sigma^2_j(\theta)a_n\right) \\
&\leq P\left(|\tilde{\sigma}^2_{n,j}(\theta)-\sigma^2_j(\theta)| > 3\sigma^2_j(\theta)a_n\right),
\end{aligned}$$

where all relationships are elementary and we have used that $a_n = o(1)$ implies that there is n large enough s.t. $2a_n + a_n^2 < 3a_n$ and $-2a_n + a_n^2 > -3a_n$. From here we conclude that

$$\begin{aligned}
P(|\tilde{\sigma}_{n,j}(\theta)-\sigma_j(\theta)| > \sigma_j(\theta)a_n)(3\sigma_j(\theta)a_n)^{1+\psi/2} &\leq P(|\tilde{\sigma}^2_{n,j}(\theta)-\sigma^2_j(\theta)| > 3\sigma^2_j(\theta)a_n)(3\sigma_j(\theta)a_n)^{1+\psi/2} \\
&\leq E(|\tilde{\sigma}^2_{n,j}(\theta)-\sigma^2_j(\theta)|^{1+\psi/2}) \\
&= E\left(\left|\frac{\sum_{i=1}^n(m_j(Z_i,\theta)-E[m_j(Z_i,\theta)])^2-\sigma^2_j(\theta)}{n}\right|^{1+\psi/2}\right) \\
&= n^{-\psi/2}E\left(\frac{|\sum_{i=1}^n(m_j(Z_i,\theta)-E[m_j(Z_i,\theta)])^2-\sigma^2_j(\theta)|^{1+\psi/2}}{n}\right) \\
&\leq n^{-\psi/2}E\left(\frac{\sum_{i=1}^n|(m_j(Z_i,\theta)-E[m_j(Z_i,\theta)])^2-\sigma^2_j(\theta)|^{1+\psi/2}}{n}\right) \\
&= n^{-\psi/2}E\left[|(m_j(Z_i,\theta)-E[m_j(Z_i,\theta)])^2-\sigma^2_j(\theta)|^{1+\psi/2}\right],
\end{aligned}$$

where we have used Markov's and Holder's inequalities. $\square$

Lemma A.6. *Assume Assumption A.1 and that $\theta \in \Theta$ satisfies $E[||m(Z,\theta)||^{2+\psi}] < \infty$ for some $\psi > 0$. For all $j = 1,\ldots,J$ and any sequence $\{a_n\}_{n\geq 1} = o(1)$,*

$$\begin{aligned}
&P(|\tilde{\sigma}_{n,j}(\theta)-\hat{\sigma}_{n,j}(\theta)| > a_n)(3\sigma_j(\theta)a_n)^{1+\psi/2} \\
&\leq n^{-1-\psi}E[|m_j(Z_i,\theta)-E[m_j(Z_i,\theta)]|^{2+\psi}] + n^{-\psi/2}E[|(m_j(Z_i,\theta)-E[m_j(Z_i,\theta)])^2-\sigma^2_j(\theta)|^{1+\psi/2}].
\end{aligned}$$

Proof. First, consider the following argument:

$$
\begin{aligned}
&P\left(|\tilde{\sigma}_{n,j}(\theta)-\hat{\sigma}_{n,j}(\theta)|>a_n\right)\left(3\sigma_j(\theta)a_n\right)^{1+\psi/2}\\
&=P(|\tilde{\sigma}^2_{n,j}(\theta)-\hat{\sigma}^2_{n,j}(\theta)|^{1+\psi/2}>(3\sigma_j(\theta)a_n)^{1+\psi/2})(3\sigma_j(\theta)a_n)^{1+\psi/2}\\
&\leq E(|\tilde{\sigma}^2_{n,j}(\theta)-\hat{\sigma}^2_{n,j}(\theta)|^{1+\psi/2})\\
&=E\left(\left|\frac{\sum_{i=1}^n[(m_j(Z_i,\theta)-\bar{m}_{n,j}(\theta))^2-(m_j(Z_i,\theta)-E[m_j(Z_i,\theta)])^2]}{n}\right|^{1+\psi/2}\right)\\
&=E(|\bar{m}_{n,j}(\theta)-E[m_j(Z_i,\theta)]|^{2+\psi})\leq n^{-1-\psi}E(|m_j(Z_i,\theta)-E[m_j(Z_i,\theta)]|^{2+\psi}),
\end{aligned}
$$

where every relationship is elementary. Now, consider the following derivation:

$$
\begin{aligned}
&P(|\tilde{\sigma}_{n,j}(\theta)-\hat{\sigma}_{n,j}(\theta)|>a_n)\\
&=P(\{\hat{\sigma}_{n,j}(\theta)>a_n+\tilde{\sigma}_{n,j}(\theta)\}\cup\{\hat{\sigma}_{n,j}(\theta)<-a_n+\tilde{\sigma}_{n,j}(\theta)\})\\
&=P(\{\hat{\sigma}^2_{n,j}(\theta)-\tilde{\sigma}^2_{n,j}(\theta)>a_n^2+2a_n\tilde{\sigma}_{n,j}(\theta)\}\cup\{\hat{\sigma}^2_{n,j}(\theta)-\tilde{\sigma}^2_{n,j}(\theta)<a_n^2-2a_n\tilde{\sigma}_{n,j}(\theta)\})\\
&\leq\left\{\begin{array}{c}P(\{\hat{\sigma}^2_{n,j}(\theta)-\tilde{\sigma}^2_{n,j}(\theta)>a_n^2(1+2\sigma_j(\theta))+2a_n\sigma_j(\theta)\}\cup\{\hat{\sigma}^2_{n,j}(\theta)-\tilde{\sigma}^2_{n,j}(\theta)<a_n^2a_n^2(1+2\sigma_j(\theta))-2a_n\sigma_j(\theta)\})\\+P(|\tilde{\sigma}_{n,j}(\theta)-\hat{\sigma}_{n,j}(\theta)|>\sigma_j(\theta)a_n)\end{array}\right\}\\
&\leq P(|\hat{\sigma}^2_{n,j}(\theta)-\tilde{\sigma}^2_{n,j}(\theta)|>3a_n^2\sigma_j(\theta))+P(|\tilde{\sigma}_{n,j}(\theta)-\hat{\sigma}_{n,j}(\theta)|>\sigma_j(\theta)a_n),
\end{aligned}
$$

where all relationships are elementary and we have used that $a_n=o(1)$ implies that there is n large enough s.t. $a_n^2(1+2\sigma_j(\theta))+2a_n\sigma_j(\theta)<3a_n\sigma_j(\theta)$ and $a_n^2(1+2\sigma_j(\theta))-2a_n\sigma_j(\theta)>-3a_n\sigma_j(\theta)$. From here we conclude that

$$
\begin{aligned}
&P(|\tilde{\sigma}_{n,j}(\theta)-\hat{\sigma}_{n,j}(\theta)|>a_n)(3\sigma_j(\theta)a_n)^{1+\psi/2}\\
&\leq P(|\hat{\sigma}^2_{n,j}(\theta)-\hat{\sigma}^2_{n,j}(\theta)|>3a_n\sigma_j(\theta))(3\sigma_j(\theta)a_n)^{1+\psi/2}+P(|\tilde{\sigma}_{n,j}(\theta)-\sigma_j(\theta)|>\sigma_j(\theta)a_n)(3\sigma_j(\theta)a_n)^{1+\psi/2}\\
&\leq n^{-1-\psi}E(|m_j(Z_i,\theta)-E[m_j(Z_i,\theta)]|^{2+\psi})+n^{-\psi/2}E[|(m_j(Z_i,\theta)-E[m_j(Z_i,\theta)])^2-\sigma_j^2(\theta)|^{1+\psi/2}],
\end{aligned}
$$

where we have used Lemma A.5. □

Lemma A.7. *Assume Assumption A.1 and that $\theta\in\Theta$ satisfies $E[||m(Z,\theta)||^{2+\psi}]<\infty$ for some $\psi>0$. For all $j=1,\ldots,J$ and any sequence $\{a_n\}_{n\geq1}=o(1)$,*

$$
\begin{aligned}
&P(|\hat{\sigma}_{n,j}(\theta)-\sigma_j(\theta)|>2a_n)(3\sigma_j(\theta)a_n)^{1+\psi/2}\\
&\leq n^{-1-\psi}E[|m_j(Z_i,\theta)-E[m_j(Z_i,\theta)]|^{2+\psi}]+n^{-\psi/2}2E[|(m_j(Z_i,\theta)-E[m_j(Z_i,\theta)])^2-\sigma_j^2(\theta)|^{1+\psi/2}].
\end{aligned}
$$

Proof. By triangular inequality, $|\hat{\sigma}_{n,j}(\theta)-\sigma_j(\theta)|\leq|\hat{\sigma}_{n,j}(\theta)-\tilde{\sigma}_{n,j}(\theta)|+|\tilde{\sigma}_{n,j}(\theta)-\sigma_j(\theta)|$ and therefore

$$
\begin{aligned}
P(|\hat{\sigma}_{n,j}(\theta)-\sigma_j(\theta)|>2a_n)&\leq P(\{|\hat{\sigma}_{n,j}(\theta)-\tilde{\sigma}_{n,j}(\theta)|>a_n\}\cup\{|\tilde{\sigma}_{n,j}(\theta)-\sigma_j(\theta)|>a_n\})\\
&\leq P(|\hat{\sigma}_{n,j}(\theta)-\tilde{\sigma}_{n,j}(\theta)|>a_n)+P(|\tilde{\sigma}_{n,j}(\theta)-\sigma_j(\theta)|>a_n).
\end{aligned}
$$

The result follows from the previous inequality and Lemmas A.5 and A.6 . □

Lemma A.8. *Assume Assumption A.1 and that $\theta\in\Theta$ satisfies $E[||m(Z,\theta)||^{2+\psi}]<\infty$ for some $\psi>0$. For all $j=1,\ldots,J$ and any $c>0$,*

$$P(|\hat{\sigma}_{n,j}(\theta)-\sigma_j(\theta)>n^{-1/2-c})=O(n^{1/2-\psi/4+(1+\psi/2)c}).$$

Under Assumption $A^\star.5$ the RHS expression is $o(n^{-1/2})$ for a choice of $c>0$ that is small enough.

Proof. Let $c\in(0,1/2)$. By Lemmas A.1 and A.7 applied to $a_n=n^{-1/2-c}/2=o(1)$, we conclude that

$$P(|\hat{\sigma}_{n,j}(\theta)-\sigma_j(\theta)|>n^{-1/2-c})=(3\sigma n^{-1/2-c}/2)^{-1-\psi/2}O(n^{-1-\psi}+n^{-\psi/2})=O(n^{1/2-\psi/4+(1+\psi/2)c}).$$

To complete the proof, we notice that $O(n^{1/2-\psi/4+(1+\psi/2)c}) = o(n^{-1/2})$ if and only if $\psi > (4+2\psi)c+4$. By making c arbitrarily small, the condition can be achieved whenever $\psi > 4$, i.e., under Assumption A*.5. $\square$

References

Bugni, F. A. (2014): "A comparison of inferential methods in partially identified models in terms of error in coverage probability (Formerly circulated as "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Elements of the Identified Set")," Accepted for publication in *Econometric Theory* on August 11, 2014.

Joag-Dev, K. and F. Proschan (1983): "Negative association of random variables with applications," *The Annals of Statistics*, 11, 286–295.